PROTOPHENOMENA AND THEIR NEURODYNAMICAL CORRELATES

Bruce MacLennan,*
CS-96-331

July 18, 1996

Abstract

The 'hard problem' is hard because of the special epistemological status of consciousness, which does not, however, preclude its scientific investigation. Data from phenomenologically trained observers can be combined with neurological investigations to establish the relation between experience and neurodynamics. Although experience cannot be reduced to physical phenomena, parallel phenomenological and neurological analyses allow the structure of experience to be related to the structure of the brain.

Such an analysis suggests a theoretical entity, an elementary unit of experience, the *protophenomenon*, which corresponds to an activity site (such as a synapse) in the brain. The structure of experience is determined by connections (e.g. dendrites) between these activity sites; the connections correspond to temporal patterns among the elementary units of experience, which can be expressed mathematically. This theoretical framework illuminates several issues, including degrees of consciousness, nonbiological consciousness, sensory inversions, unity of consciousness and the unconscious mind.

*I am grateful to David Chalmers, Jonathan Shear and two anonymous referees for many helpful criticisms and suggestions on two previous drafts of this report, which is an expanded version of MacLennan (in press).

1 Why the 'Hard Problem' is Hard

1.1 Special Epistemological Status of Consciousness

I take the 'hard problem' of consciousness to be to understand the relation between our subjective experience and the brain processes that cause it; that is, to reconcile our everyday feeling of consciousness with the scientific worldview (MacLennan, 1995). This problem is hard because consciousness has unique epistemological characteristics, which must be accommodated by any attempted solution. I will summarize these characteristics; more detail can be found in Searle (1992, chs. 4, 5) and Chalmers (1995, 1996), whose positions, if I have understood them correctly, are consistent with mine.[1]

First, science is a public enterprise; it attains knowledge that is independent of the individual investigator by limiting itself to public phenomena. Ultimately it is grounded in shared experiences, for example, when we both look at a thermometer and read the same temperature. Traditionally science has accomplished its ends by focusing on the more public, objective aspects of phenomena (e.g. temperature as measured by a thermometer), and by ignoring the more private, subjective aspects (how warm it feels to me). In other words, science has restricted itself to facts about which it is easy to reach agreement among a consensus of trained observers. Although this restriction has aided scientific progress, it prevents the scientific study of consciousness, which is essentially private and subjective.[2]

Second, science's neglect of the subjective is also apparent in its reductive methods. For example, once the experiential phenomenon of temperature has been separated into its subjective and objective parts (felt vs. measured temperature), the objective part can be reduced to other objective phenomena (mean kinetic energy of molecules), but the subjective components of the original phenomenon remain unreduced. Although this approach has been very fruitful for the development of physical theory, it fails when the topic of the investigation is precisely that subjectivity that it ignores.

[1]A more detailed comparison will be found at the end of this paper.

[2]It should be apparent that I am using 'subjective' and 'objective' to distinguish private, 'first person' phenomena from public, 'third person' phenomena. As Searle (1992) observes, progress on the mind-body problem has been impeded by the connotations acquired by these terms, viz., the objective is unbiased and factual, whereas the subjective is biased or distorted. Indeed, I will argue for the possibility of unbiased, factual statements about subjective (private, first person) phenomena.

In summary, the standard reduction pattern in science, which reduces the objective to the objective, cannot solve the hard problem, which deals with the relation between the subjective and the objective. If reduction is to play a role at all, it must take a different form.

Finally, science traditionally seeks facts — observations — that are independent of the observer; this supposes that the observer can be separated from the observed (another aspect of the subject-object distinction). However, in confronting the hard problem we cannot separate the observer and the observed, for consciousness is observation, the subject experiencing the object. That is, experience comprises both observer and observed, the termini of the vector of consciousness. Separating the two breaks the very connection that we aim to study.

1.2 Scientific Investigation of Consciousness

The preceding observations might suggest that the hard problem is invulnerable to scientific methods, but I believe that progress may be made by loosening a few of science's self-imposed restrictions, many of which are relics of long discredited philosophies of science, such as naive empiricism and logical positivism. Consciousness is our opening to the world; it is the vehicle by which we experience anything. Therefore we cannot *observe* consciousness per se, since we observe *through* consciousness. Nevertheless, with practice we can identify characteristics of consciousness that are relatively independent of its content, and in this way separate them from its content.

An analogy may make this clear. The aperture of a camera is its 'window to the world', since any image in the camera must come through the aperture. (For the sake of the analogy we suppose the camera cannot be opened in any way.) From within the camera the aperture per se is not visible; all we can see is the image it transmits, the scene at which it is aimed. Although the aperture is visible only by virtue of the images it transmits, observation nevertheless shows that certain characteristics of the image (focus, brightness, depth of field) are more a consequence of the aperture than of its content. Thus the aperture may be investigated indirectly. So also we may investigate the structure of consciousness independently of its content.

It may seem that by advocating such private 'observation' of consciousness, we have abandoned all hope of publicly validatible science, but it is worth remembering that all observation is ultimately private. Science has developed methods (such as measurement) that, in a context of shared training and experience, lead to general agreement among qualified observers (with varying theoretical commitments), and thus provide a reasonably stable body of public facts, which may be used for the support or critique of theories.[3] To bring consciousness into the scope of science will require a body of appropriately trained observers; the public facts necessary for a scientific theory of consciousness will emerge from their consensus.

The camera analogy shows the importance of training, for the relevant phenomena, e.g. depth of field, might not be apparent to untutored observers. The difficulties with 'split-brain' and 'blind-sight' patients as informants also illustrate the need for trained observers. I believe that the best example of the kind of training required comes from phenomenological philosophy and psychology (see, e.g., Ihde, 1986).

In summary, although consciousness cannot be reduced to physical phenomena by the standard reductive methods of the sciences, it can be investigated to yield publicly validatible facts about the structure of consciousness, which can be related, in turn, to the observations of neuroscience.

1.3 Phenomenology

Phenomenology studies the structure of *phenomenal worlds*, that is, the worlds actually experienced by individuals. Henceforth 'phenomenon' will be used in a technical sense: a *phenomenon* is anything that appears in consciousness, anything we experience, no matter what its origin. For example, perceptions, recollections, dreams, pains (whether real or phantom), mental images, mental dialogues, moods, anticipations, desires and hallucinations are some of the kinds of phenomena. Further, your phenomenal world determines the structure of possible phenomena, and the *state* of your phenomenal world at a time is equivalent to the content of your consciousness at that time. That is, your phenomenal world is a structure of *potential*

[3] One cannot ignore the importance of training, shared experience and institutions in the creation of 'facts'. Even something so simple as accurately reading a thermometer requires training and skill (e.g. reading the top or bottom of the meniscus). Training is all the more necessary for reading bubble-chamber images and gas chromatographs. The histories of N-rays and polywater show how competent observers can disagree over even the existence of a phenomenon (let alone its measurement); 'cold fusion' is a more recent example. See Fleck (1979) for an informative case study.

experiences; at any given time one of these is actualized as your conscious experience at that time.

The phenomena are the starting point of all science, for they are what is given to us (cf. Latin *data* = given things).[4] However, this is easily misunderstood for, at least since the appearance of logical positivism in the philosophy of science, there has been a tendency to suppose that the phenomena are simple things, such as sense data. 'Red-here-now', that is, the current experiencing of a patch of red at a particular location in the visual field, is a well-known example. The phenomenologists, especially Husserl and Heidegger, have demonstrated the incorrectness of this view, for rarely, if ever, do we actually experience red-here-now; they have revealed some of the complexity of real phenomena.

Suppose, for example, you rotate an ordinary die in front of me and ask for a phenomenological account of what I see.[5] I would be incorrect to describe a certain arrangement of black ovals in white parallelograms, both of systematically changing shape. That does not accurately describe the phenomenon as I experience it, for I recognize the object and so it is seen as a die, and I see it rotating in space, not changing shape in some mysterious way. Even if I were unfamiliar with dice, I would see the rotation of a white cube marked with spots. Indeed, it would take very unusual conditions to make me see the die as parallelograms and ovals changing shape. (Such a situation, a consequence of a brain tumor, is described by Oliver Sacks in the title essay of his *Man who Mistook his Wife for a Hat*.)

An additional complexity of phenomena is that they are not entirely in the here-and-now; for example, my current experience of the die includes some foreshadowing of future possible experiences. Thus we have expectations — some vague and others precise in accord with our familiarity with dice — about what we will see as the die is rotated. These expectations go beyond the visual; for example, we also have expectations about the hardness and weight of dice, and if we see two dice in an open palm, we have the expectation of some kind of dice game. All these and more are part of 'phenomenal field' surrounding the visual perception of the die. Further, we see that much of the

[4]This is true of the empirical sciences, but also of the so-called a priori sciences, such as mathematics, which start from the apparently invariable structure of the phenomenal world.

[5]The die example derives from Husserl's *Cartesian Meditations* (§§17–19), where it is developed at length.

phenomenon is a construct, both of the culture and of individual experience.

We must include expectation as part of our current conscious experience because, for example, we are in a different conscious state if we come to the pantry door expecting the shelves to be bare or come expecting them to be full. To the extent that the expectations contained in a phenomenon are met, we experience normality and familiarity, to the extent they are not, we experience dissonance and novelty. Phenomena appear gradually over the horizon of consciousness, and as they do so they actualize some, but not all, of the possibilities that may have been foreshadowed in the current state. This gradual actualization of foreshadowed phenomena creates the continuity of subjective time.[6]

In summary, phenomena are not simple; they are highly complex and strongly coupled to the rest of the phenomenal world in which they appear. Therefore some training is necessary to be able to observe phenomena accurately and to analyze the structure of the phenomenal world. Nevertheless I believe that phenomenological training of this kind provides a basis for making the characteristics (though not the experience) of consciousness public.[7]

2 Protophenomena

2.1 Decomposition of Phenomenal Worlds

I have argued that the subjective is not reducible to the objective according to the usual pattern in science. Nevertheless it is important to strive for some kind of reduction of the more complex to the simpler or better understood. This can be accomplished by an analysis of the structure of consciousness, which allows a phenomenological subjective-to-subjective reduction that parallels a neurological objective-to-objective reduction. It is to be expected that progress in each reduction will facilitate the other, in turn. In this section I'll outline the results of such a process, which suggests a theoretical entity that may be useful in constructing a scientific theory of consciousness.

[6] Pribram (1991, 214–220) argues that the alternations of familiarity and novelty parse experience into *episodes*.

[7] Despite the importance of phenomenology, in this essay I have avoided its technical terminology, which would be more confusing than helpful.

At the highest level the phenomenal world can be analyzed along modal and functional lines (appearance, sound, smell, memory, intention, etc.), but the rotating-die example shows that these components are far from independent (the rotating die phenomenon is not visual alone, but includes kinesthetic and other aspects). Fortunately we can expect neuroscientific investigations of functional areas and pathways will correct erroneous preconceived ideas about the structure of the phenomenal world. The resulting analysis of consciousness into components of different kinds can be called a *qualitative* reduction.

A different kind of reduction, which analyzes some aspect of consciousness into constituents of a like kind, may be called *quantitative*. This analysis is suggested by *topographic maps*, which are ubiquitous in the brain. A familiar example is the *somatotopic map* in the somatosensory cortex: nearby parts of the body are mapped to nearby parts of the cortex, so that the arrangement of neurons mimics the arrangement of the body. Similarly, in the early vision areas we find *retinotopic maps*, where neurons are arranged in a pattern mimicking the arrangement of their receptive fields in the retina.

The *receptive field* of a neuron in a topographic map refers to the stimuli to which it responds; for example, a neuron in a somatosensory map might respond to pressure on a particular patch of skin, or a neuron in the visual cortex to light on a particular patch of the retina. In such a case we can identify the subjective experience corresponding to activity in this neuron, namely, the feeling of pressure in that patch of skin, or the sensation of light on that patch of retina. I call such a 'little bit of experience' a *phenomeniscon* or *protophenomenon*.[8] Further, we can see how, to a first approximation, the objective neurological processes corresponding to tactile or visual sensation can be reduced to a large number of receptive fields of this kind. This suggests phenomenological subjective-to-subjective reductions (phenomena to protophenomena) paralleling neurological objective-to-objective reductions (e.g., topographic maps to their neurons, and sensory surfaces of skin to receptive fields).

This may seem to be a return to the red-here-now model of phenomena, but there are important differences. First, neurons have *functional recep-*

[8]This is an approximate definition; protophenomena are described more precisely in the following sections. 'Phenomeniscon' (accent on penult), a diminutive of 'phenomenon', is used in MacLennan (1995). I am grateful to David Chalmers for suggesting the alternative 'protophenomenon'.

tive fields that are more abstract than simple spatial patches. For example, primary visual cortex contains neurons whose receptive fields are four-dimensional combinations of retinal location, spatial orientation and spatial frequency (see MacLennan, 1991, for a survey). Indeed, Pribram (1991, 79–83) has stressed that they are not limited to these four dimensions, but respond to many additional dimensions of the stimulus.

Next, as shown by the rotating-die example, there is much more to ordinary phenomena than the sense data, so we will have to take account of nonsensory protophenomena that represent the constituents of expectations, interpretations, intentions and many more abstract properties. Third, few neurons have simple fixed receptive fields, since even sensory neurons receive inputs from higher brain areas; therefore, virtually all protophenomena depend on other protophenomena. Finally, although I have used a simple sensation as an example, because its protophenomena are easy to visualize, sensation is only one aspect of most phenomena (many of which involve no sensation).

The preceding model can be extended to nonsensory neurons as follows. The activity of a sensory neuron reflects the presence of a stimulus in its receptive field, which is a region of some (possibly abstract) physical space (pressure, light, sound, heat, space, frequency, orientation, etc.); it responds maximally to stimuli in that region. Similarly, the activity of a nonsensory neuron reflects the activities of other neurons, and so it has a *functional receptive field*, comprising certain patterns of activities of its input neurons, to which it responds. That is, as sensory neurons respond to combinations of physical energy, so nonsensory neurons respond to combinations of neural activity. Turning to the subjective side, we see that, as a sensory protophenomenon corresponds to the experience of the physical phenomena in a sensory neuron's receptive field, so a nonsensory protophenomenon corresponds to the experience of combinations of other protophenomenal intensities, those intensities corresponding to activity in the nonsensory neuron's functional receptive field.[9]

[9]This description is simplified for the sake of exposition, since sensory neurons also respond to other neurons, and so their complete receptive field includes the activities of other neurons. Correspondingly, the subjective intensity of a sensory protophenomenon depends on the subjective intensities of other protophenomena as well as on the experience of objective physical processes.

2.2 Definition of Protophenomena

Chalmers (1995) has argued that 'a theory of consciousness requires the addition of *something* fundamental to our ontology' and that 'where there is a fundamental property, there are fundamental laws'. In this section I will propose protophenomena as a candidate for this 'something' and describe fundamental laws governing them (analyzed mathematically in the Appendix).

We have seen that activity in a neuron reflects the extent to which its functional receptive field is occupied at that point in time. Subjectively, this activity corresponds to the intensity in consciousness of a protophenomenon corresponding to the site of that activity. Therefore I hypothesize a one-to-one relationship between protophenomena and certain *activity sites* in the brain, and further hypothesize that the intensity of a protophenomenon varies directly with the neurological activity at that site. What are these activity sites?

Following Sherrington, who said, 'Reflex action and mind seem almost mutually exclusive — the more reflex the reflex, the less does mind accompany it', Pribram has argued that consciousness is associated with graded dendritic processes rather than all-or-nothing axonal spiking (Miller, Galanter & Pribram, 1960, pp. 23–4; Pribram, 1971, pp. 104–5; Pribram, 1991, pp. 7–8). For concreteness I will accept this hypothesis and take synapses to be the activity sites (though the identification is not crucial to most of the following).[10] Candidates for the activity of the synapse include presynaptic membrane potential, postsynaptic membrane potential and neurotransmitter flux across the synaptic cleft. Since for the most part each is proportional to the others, it doesn't matter much which we pick; for concreteness, I'll hypothesize postsynaptic membrane potential.

The easiest way to understand protophenomena is to think of them as the atoms (indivisible constituents) of consciousness. As atoms make up macroscopic objects (trees, tables, etc.), so protophenomena make up phenomena. In both cases, the effect of each individual element on the whole is usually minute. Indeed, our phenomenal world comprises perhaps 10^{14} to 10^{15} protophenomena (i.e., the number of synapses in a human brain). Normally a change in the intensity of a single protophenomenon will not be perceived since it will not usually lead to a macroscopic change in conscious

[10] In particular, my proposal does not exclude the possibility that microtubules, as proposed by Hameroff (1994), are among the activity sites.

state, such as a judgement. Normally the intensity of a large number of prototphenomena must change in a coherent way for a pehnomenon to appear in consciousness, that is, for there to be a macroscopic change in conscious state.

Further, just as the physical world is more than the sum-total of the macroscopic objects it contains (for there are are lots of particles other than those bound in macroscopic objects), so also conscious experience is more than the sum-total of the (macroscopic) phenomena that appear in it (for conscious experience contains many protophenomena that are not part of any macroscopic phenomenon).

Hence, every protophenomenon is experienced — by definition, since protophenomena are defined to be the elementary constituents of experience. However, differences in a single protophenomenon's intensity will not usually be perceived, in the following sense: a change in an individual protophenomenon's intensity will not usually lead to a macroscopic change in conscious state [i.e. lead to a change in the (macroscopic) phenomena]. For example, increasing the intensity of a single protophenomenon will not normally cause you to think 'I see it' instead of thinking 'I don't see it.' Analgously, changing the motion of a single atom will not normally change the macroscopic state of a physical system (e.g., temperature, viscosity, shape, rigid motion). In terms of microscopic state, however, changing a single protophenomenon changes the conscious state (in this sense all protophenomena are experienced), just as changing a single atom changes the physical state. This example can be made quite concrete. If I change the activity in one of my 100 million retinal receptor cells, it will make a (very tiny) change in my conscious state, but it will not normally cause a macroscopic change in conscious state (e.g., leading me to note the change to myself or to mention it to someone else).

Therefore, the Hard Problem does not reappear in the relation of protophenomena to phenomena (e.g., 'How do unexperienced protophenomena add up to experience?') All protophenomena are experienced in the sense that a change in protophenomenal intensity is equivalent to a change in conscious state. Or, we may say that all protophenomena have the quality of subjectivity, so we don't have the problem of how macroscopic subjectivity emerges from microscopic nonsubjectivity; rather, we are dealing with the emergence of macroscopic subjectivity from microscopic subjectivity (which is no more problematic than the emergence of temperature from molecular motion).

On this view, the state of the phenomenal world, that is, the content of consciousness, is identical with the intensities of all the protophenomena. The appearance of coherent or stable phenomena can be identified with cohesive or coherent patterns of intensity among the protophenomena (just as macroscopic objects and events can be identified with cohesive and coherent patterns of activity among atoms).

An interesting question is just how the intensities of protophenomena combine to form a conscious state. A recent analysis by Sanger (submitted) of activity in populations of neurons suggests an answer. The functional receptive field of a neuron is proportional to a *conditional probability density field* (CPDF) σ_i, defined over possible stimuli x, which determines the probability $\sigma_i(x)$ a given stimulus will cause the neuron to fire (more precisely, it determines the rate of a Poisson firing process). Collectively, a population responding to a common set of inputs has a CPDF that is proportional to the (pointwise) product of the fields of the neurons firing at a given time.

$$\sigma_{\text{pop}} = \prod_{i \in \text{spike}(x)} \sigma_i,$$

where $\prod$ represents a pointwise product of the fields,

$$\sigma_{\text{pop}}(x) = \prod_{i \in \text{spike}(x)} \sigma_i(x). \tag{1}$$

This analysis can be transferred to the phenomenal realm as follows. Each protophenomenon has an associated CPDF defined over the protophenomena upon which it depends (its "inputs"). (In fact, the conditional probability density of an input signal is proportional to the temporal convolution of that signal with the protophenomenon's characteristic pattern, defined in Section 2.6.) A population of protophenomena dependent on the same input protophenomena has a CPDF that is the product of the CPDFs of all the high-intensity input protophenomena, that is, of all the input protophenomena present in the current conscious state. The CPDFs of individual protophenomena can be quite broad, but in the joint response to the same input of a large number, the product can be very narrow, so that they define a phenomenal state quite precisely.

Equation 1 can be rearranged in a way that makes its information processing role clearer. Let $s_i(x) = 1$ if $i \in \text{spike}(x)$ and $s_i(x) = 0$ if $i \notin \text{spike}(x)$.

Then we can rewrite Eq. 1,

$$\sigma_{\text{pop}}(x) = \prod_i [\sigma_i(x)]^{s_i(x)}.$$

Taking logarithms eliminates the awkward products and exponents:

$$\log \sigma_{\text{pop}}(x) = \sum_i s_i(x) \log \sigma_i(x).$$

Now, $s_i(x)$ is zero or one depending on whether, over a given short interval of time, neuron i fires for stimulus x. Therefore, on average, $s_i(x)$ will be the probability that neuron i fires on stimulus x, which is just $\sigma_i(x)$. Thus we write,

$$\log \sigma_{\text{pop}}(x) \approx \sum_i \sigma_i(x) \log \sigma_i(x),$$

and we see that the right-hand side has the form of *negentropy* (negative entropy, i.e., ordered information), $\log \sigma_{\text{pop}}(x) \approx -H\{\sigma_i(x)\}$. We must be careful, however, in interpreting this quantity as entropy, since we have not normalized, $\sum_i \sigma_i(x) = 1$. Nevertheless, the formal similarity to an entropy suggests that the combination of conditional probability density fields — in either the physical or phenomenal realms — may be interpreted in an information-theoretic framework.

2.3 Ontological Status of Protophenomena

Are protophenomena real? At this time I believe it is best to treat protophenomena as *theoretical entities* (Hempel, 1965, pp. 177–9; Maxwell, 1980), that is, hypothetical constructs that are postulated for the sake of the theory, and are validated by their explanatory value and fruitfulness for scientific progress. (Quarks are examples of theoretical entities in contemporary physics.)

Here again the atomic analogy is helpful. When atoms were first postulated, they were theoretical entities; indeed it is only in recent years that they have become observable (still, of course, through instruments). At first many respected scientists denied their existence, while admitting their convenience for theory. In time, their explanatory value became so great that they were accepted as real.

Now we accept atoms (or more elementary particles) as the ultimate constituents of matter, which cause the properties of macroscopic objects, and in this sense, in physical theory, atoms are prior to trees. Nevertheless, in experience, trees are prior to atoms. Similarly, in theoretical phenomenology, protophenomena are prior to phenomena, for they are the causes of phenomena, but in experience phenomena are prior to protophenomena; we experience trees, and by analysis break the phenomena down into protophenomena.

Another ontological issue is whether an isolated synapse (in a petri dish or simple organism, for example) has an associated protophenomenon. One answer is that protophenomena, as components of the phenomenal world, make sense ('exist') only in the context of a sufficiently complex nervous system. ('Sufficiently complex' is of course a matter of degree.) Such 'emergent existence' is not uncommon in scientific theories. For example, sound is a compression wave in air or another medium. The theory assigns a pressure to every point in the medium, yet it makes little sense to talk about the pressure (or sound) of an isolated air molecule. The compression wave, which comprises elementary units of pressure assigned to individual molecules, makes sense only in the context of a large number of molecules. Similarly, I think it may make sense to assign protophenomena to activity sites only in the context of a large number of activity sites. This emergence does not make the protophenomena any less real; they are as real as the elementary units of pressure which constitute the sound wave.[11]

I have said that conscious states are the totality of protophenomenal intensities, so phenomena, as aspects of the phenomenal world, are cohesive and coherent patterns of protophenomenal intensity. I see no reason to hypostasize these patterns by postulating (subjective) entities corresponding to phenomena (images, ideas, perceptions, etc.). The coherence of the intensity patterns constitutes the appearance of macroscopic phenomena in experience (as will be explained in more detail later). By analogy, to explain the coherent physical effect of a baseball on a window it is not necessary to postulate the existence of anything beyond the baseball's constituent atoms, such as

[11]By again transferring Sanger's (submitted) analysis to protophenomena, we may say that a small number of protophenomena so weakly delimit the inputs to which they might be responding that we can hardly say there *is* a definite conscious state. Conversely, a large number of protophenomena can define the possible inputs quite precisely, so that it is useful to talk of a definite conscious state.

a 'ball entity', to represent the ball's coherence. So also, the collective action of protophenomena are sufficient to explain the experience of a coherent phenomenon.

2.4 Protophenomena Correspond One-to-one With Activity Sites

Next I must explain why I have claimed that protophenomena correspond one to one with activity sites. First, I take it as given that phenomenal differences imply neural differences, that is, that a difference in conscious state is dependent on an underlying difference in neural (or physiological) state. Denying this would permit conscious states unsupported by physical states, which would, it seems to me, undermine the whole project of reconciling conscious experience with the scientific world view, the *raison d'être* of the 'hard problem'.

Second, I hypothesize that differences in activity at activity sites imply differences in conscious experience. Here the reason is Occam's Law, for we would otherwise have to suppose that some activity sites (e.g. synapses) have associated protophenomena while others don't. Although this may be the case, I see no evidence supporting it.[12] In any case, this hypothesis is not necessary for the overall theory of protophenomena.

2.5 Structure of Phenomenal Worlds

So far I have discussed protophenomena as elementary units of experience, but I have had little so say about how they are assembled into a phenomenal world and the phenomena it reveals. Clearly, the phenomenal world is spread out in space; although it is generated predominantly in the brain, it is projected 'out there': I feel indigestion in my stomach, not in my brain; I see the approaching cars in front of me, not in my visual cortex. How are

[12]Note, that the claim is only that differences of activity lead to differences of protophenomenal intensity and hence (microscopic) changes in conscious state, not necessarily that the difference will have a significant effect on future (macroscopic) conscious states, or that it will change behavior, judgements, verbal reports or stored memory. Analogously, changing a pixel changes the picture, but such a change would not make a difference normally. The issue of the unconscious is treated in Section 3.5; suffice it to say that it does not contradict the hypothesized one-to-one relation.

neural events in the brain projected into the body and surrounding space? For example, what makes me experience activity in a certain neuron as pain in my finger and not pain in my toe?

Our discussion of topographic maps suggests that they have an important role to play, but how, precisely, do the spatial relations among neurons lead to phenomenal relations (such as perceived spatial and more abstract relations) among protophenomena? Although there may be some diffuse electrical and chemical effects on the activity of neurons, it seems that in general the spatial arrangement of neurons is significant only because it correlates with connectivity: nearer neurons are more likely to be connected than are more distant ones, and connections create dependencies between neurons.

Specifically, connections between neurons create dependencies between their activities. Thus, if one neuron synapses on another, then the activity of the first will tend to increase or decrease the activity of the second (depending on whether the synapse is excitatory or inhibitory). Also, if one neuron synapses on two others, it will indirectly establish a (positive or negative) correlation between the activities of the two postsynaptic neurons.

We have corresponding dependencies in the phenomenal realm. Increased intensity of one protophenomenon can tend to increase or decrease the intensities of other protophenomena that depend on it. In this way protophenomenal dependencies constrain the possible conscious states and their evolution through time, and thus they define the necessary structure of the phenomenal world ('necessary' in the sense that this structure is invariable so long as the connections in the nervous system remain the same).

It is these dependencies between protophenomena that gives them their meaning. By analogy, a set of pixels constitutes a picture only when combined in a certain arrangement (relations of nearness or adjacency); with a different arrangement they would be a different picture; so also, with different dependencies, a set of protophenomenal intensities would constitute a different conscious state.

2.6 Protophenomenal Dependencies

Let's consider protophenomenal relations in more detail. Neurologically, the activity at a synapse is a complex spatiotemporal integration of the activities of the synapses which connect to it. To a first approximation this process is linear, and can be described by the methods of linear system analysis

(MacLennan, 1993b), which shows there is a certain spatiotemporal pattern to which the synapse shows the maximum response. Indeed, this pattern can be used to characterize the temporal response of the synapse to any spatiotemporal signal, in so far as the synapse behaves linearly. For this reason I will call this maximum-response spatiotemporal pattern the *characteristic pattern* of the synapse.[13] (Technically, the response of a synapse is a temporal convolution of its characteristic pattern with the input signal.)

This account may be transferred directly to the phenomenal realm. Each protophenomenon has a characteristic pattern, which is the spatiotemporal pattern of intensities of its input protophenomena that will maximize its intensity. Further, its characteristic pattern determines (by convolution) the protophenomenon's time-varying intensity in response to any spatiotemporal pattern in the intensities of the protophenomena on which it depends. As a consequence we can give a mathematical theory of the dynamical relations among protophenomena (see the Appendix to this paper). The characteristic patterns may be simple, as when a protophenomenon corresponds to a conjunction or disjunction of protophenomena, or more complex, as when they respond to appearance or disappearance of protophenomena, rhythmic or other temporal patterns in protophenomena, priming or inhibition of future occurrences of protophenomena, etc.

Each protophenomenon contributes its characteristic pattern to conscious experience, with its intensity at a given moment determining the degree of the pattern's presence in that moment's experience. That is, conscious experience is given by a dynamic superposition of the characteristic patterns of the protophenomena.

Overall, the dynamical relations among protophenomena are nondeterministic. First consider a synapse formed by a sensory neuron, the activity of which depends on physical stimuli as well as on the activities of other synapses. Corresponding to this activity site in the phenomenal realm we have a protophenomenon whose intensity depends on physical processes as well as on other protophenomena. But physical processes are not part of the phenomenal world, so such a protophenomenon is phenomenologically undetermined (i.e. not fully determined by other protophenomena); the physical

[13]In physics and engineering it is commonly called the *impulse response* of the system; it corresponds (via the Laplace transform) to the *transfer function*, which describes the system dynamics in terms of its transparency to different frequencies of activity.

inputs act as independent variables in the phenomenal world. In terms of the ontology of the phenomenal world, they are causal primaries, which does not imply, however, that there are not corresponding phenomenal expectations (as the rotating-die example shows). Thus sensory protophenomena are inherently nondeterministic (i.e. undetermined in the phenomenal world).

Since nonsensory protophenomena depend only on other protophenomena, to a first approximation they can be considered deterministic; indeed their responses are defined by their characteristic patterns. This is only an approximation because even nonsensory neurons depend on non-neural processes, such as the physiology of the brain, and the physical environment of the body. Although these effects can sometimes be treated as extra, hidden inputs to the synapses, they are often nonlinear and comparatively nonspecific in their effects, so it is usually better to treat them as phenomenologically undetermined alterations of the characteristic patterns of the affected protophenomena.[14]

The view advocated here might be seen as either epiphenomenalism or parallelism, but it is not; rather, it is dual-aspect monism. That is, the basic ontological claim is that there is one fundamental "stuff," and one fundamental underlying process, but that they is observed/experienced from two mutually irreducible perspectives. More specifically, what is measured from one side as activity in an activity site, is experienced from the other side as protophenomenal intensity. Indeed, according to the theory, they are numerically equal.

Causal relations in the phenomenal world parallel causal relations in the physical world, but, just as the phenomenal and physical worlds are alternative perspectives on one reality, so also phenomenal causation and physical causation are two alternative and equally valid descriptions of the constraints on the evolution of events in this reality. We can switch between the two kinds of causation as convenient for the problem at hand. The account of causation we give from the physical perspective is complete so far as the physical world is concerned. However, it cannot even talk about phenomenal processes, since they are not part of the physical world (the root of the Hard Problem). On the other hand, we can give an account of causation from the phenomenal perspective, which can explain constraints on the evolution of

[14]Thus there may be phenomenologically causeless change to the phenomenal world; an extreme example is a stroke.

the conscious state. However, the phenomenal account is less comprehensive than the physical account, since there are phenomenal events that do not depend on antecedent phenomenal events (e.g. sensation). There can be no contradiction between the phenomenological and physical accounts because they are both bound to conform to the same reality.

The foregoing sounds like epiphenomenalism, but I don't think it is in the usual sense, i.e., that consciousness is something secondary, caused by physical processes. First, since phenomenal and physical events are two perspectives on the same process, it is incorrect to speak of a causal relation between them. (Analogously, it's misleading to say the wave aspect of light causes the particle aspect, or vice versa.) Specifically, it is incorrect to ask whether activity in activity sites causes protophenomenal intensity, or vice versa, since they are two perspectives on the same thing. Therefore, we can switch between the two perspectives as convenient, and depending on whether we are more interested in the physical or phenomenal aspects of the process.

Further, I don't think it is correct to describe this theory as parallelism in the usual sense, i.e., physical and phenomenal processes that follow their own laws, do not interact, but "magically" stay in correspondence. This is because there is only one underlying "stuff," and only one underling process, and so there can be no issue concerning the interaction of two stuffs or the coordination of two processes. There is nothing mysterious about the parallelism between the physical and the phenomenal, because they are two aspects of the same thing.

* * *

I have described the protophenomenal dependencies from a mechanistic perspective; now it is worthwhile to say a few words from a functional perspective. Topographic maps show us how receptive fields are ordered in space, frequency, speed, color, and many other dimensions, and hence how their protophenomena are ordered in corresponding subjective domains. Thus dependencies among protophenomena correspond to order in a variety of dimensions. This order means that objects extended in space or other dimensions will lead to high intensities among closely dependent protophenomena (which will therefore cohere as full-fledged phenomena).

Furthermore, since change is generally continuous, or if discontinuous in some dimensions, then continuous in others, it follows that changing objects tend to move from the receptive fields they occupy to others that overlap along one or more dimensions. Think of a visual image of a moving object: it moves between overlapping spatial receptive fields; further, its edges change orientation continuously, and the light it reflects changes continuously, and so it also moves gradually from receptive field to receptive field in these dimensions. Phenomenologically, we can say that change tends to be between protophenomena that are strongly connected. Conversely, the presence (high intensity) of a protophenomenon is correlated with the future presence of the other protophenomena that depend on it. In approximate terms, the dependencies among protophenomena correspond to the likelihood (or unlikelihood) of change between protophenomena. More accurately, the characteristic pattern of a protophenomenon represents likely (excitatory) or unlikely (inhibitory) antecedent spatiotemporal patterns of protophenomena.

2.7 Phenomenological Plasticity

I have treated the phenomenal world, the structure of possible conscious states, as fixed, but it is now time to say a few words about plasticity.[15] One type of plasticity is short- or long-term change in synaptic efficacy as a result of learning or habituation, which changes the strength of the dependencies between protophenomena, or, more accurately, changes the characteristic patterns that define their time-varying intensity. These changes affect the topology (the abstract relations of near and far) and the protophenomenal dependencies of the phenomenal world. These changes, in turn, affect coherence and cohesion among protophenomena, and thus the emergence of coherent, high-level phenomena.

Second, although the adult brain does not generate new neurons, it does generate new synapses for a number of reasons, including injury and learning (Shepherd, 1994, pp. 222–3). Since I have hypothesized that protophenomena correspond one-to-one with synapses, the generation of new synapses implies the generation of new protophenomena, that is, new degrees of freedom in the phenomenal world — literally, 'expanded consciousness'. Thus we see that the phenomenal world has a flexible ontology at both the macroscopic (phenomenal) and microscopic (protophenomenal) levels.

[15] My concern here is not so much plasticity in the developing animal as plasticity in the adult.

3 Implications

I will consider briefly the implications of this theory for several issues pertaining to consciousness.

3.1 Degrees of Consciousness

By hypothesis the degrees of freedom of a phenomenal world correspond to the protophenomena it comprises, which are equal in number to the activity sites in the nervous system; further, the structural relations of the phenomenal world correspond to the connections between activity sites. Therefore, with decreasing nervous system complexity we expect a proportional decrease in both the dimension and structure of the corresponding phenomenal world. The conclusion to be drawn is that consciousness is a matter of degree; in general terms we can say that the consciousness of simpler animals is less than ours in both dimension and structure.[16]

3.2 Nonbiological Consciousness

I'll consider briefly whether the theory of protophenomena sheds any light on the perennial question of computer consciousness. From the perspective of the theory, the central question is: what sorts of physical processes have associated protophenomena, and how could we tell? For example, liquidity follows from certain physical properties of H_2O molecules, but other substances besides water may be liquid because they share these properties. Analogously, consciousness follows from certain physical properties of synapses, but other things besides brains may be conscious if their parts share these properties. Thus we need to determine sufficient conditions for the existence of protophenomena, that is, the properties of synapses (or whatever the activity sites may be) that cause them to have protophenomena.

It is possible, at least in principle, to attack this problem empirically. We would have to identify some observable protophenomenon, the presence or

[16]Chalmers (1995) tentatively reaches the same conclusion on the basis of his 'double-aspect principle'. Also, from the probabilistic perspective, for a given sensory bandwidth, the conditional probability density fields associated with the protophenomena of a simpler nervous system less sharply delimit the input phenomena; thus, for simpler nervous systems, perceptual experience is less definite.

absence (high or low intensity) of which can be reported reliably by a trained observer, and for which the corresponding synapse (or other activity site) can be identified and made accessible (e.g. through brain surgery). With care we may control some of the variables (e.g. postsynaptic potential) independently of the others (e.g. presynaptic potential), and thus determine which affect protophenomenal intensity. Indeed, one could replace the synapse by devices that are functionally equivalent in one way or another (e.g. electrically or chemically), to determine which are necessary or sufficient for the existence of the protophenomenon.

It will be objected that the investigation depends on subject report, which is a form of behavior, and therefore need not reflect subjective experience. That is correct. Since subjectivity is private, the only way such doubts can be eliminated is for the doubter to be the subject of the experiment.[17] Practically, though, the observations would become public through a consensus of trained observers of differing commitments.

From such a demonstration of protophenomena associated with nonsynaptic or even nonbiological objects we could reasonably conclude that a phenomenal world, and therefore consciousness, would emerge from sufficiently complex interconnections of those objects. Indeed, my guess is that we will find that the representational and information processing properties of synapses are all that matters, and that other physical systems with the same capabilities (such as appropriately structured massively parallel analog computers) will have protophenomena and be conscious.[18]

3.3 Origin of Sensory Qualities

A traditional conundrum in discussions of consciousness is the problem of sensory inversions (e.g., Dennett, 1991, pp. 389–98), which goes back at least to Locke: Could you tell, for example, if you experience the color spectrum oppositely from me? I believe that an improved understanding of protophe-

[17]If this seems far-fetched, it is worth noting that William McDougall requested, if he should become incurably ill, that Sherrington would perform a cerebral commissurotomy (split-brain operation) on him, so that he might directly experience its effect on his consciousness (Gregory, 1987, p. 741).

[18]It will be apparent from this that I do not accept Searle's reply to the Virtual Minds version of the System Reply to the Chinese Room Argument (MacLennan, 1993a, 1994). See also the thought experiment in Chalmers (1995).

nomenal dependencies will show that these inversions are in fact impossible, and so there is no problem to solve.[19] This is because the topology of a phenomenal space (its relations of distance and nearness) is determined by the interdependencies of its constituent protophenomena. I will use hearing to illustrate the method.

Consider first a hypothetical inversion of loud and soft sounds: Is is possible that I hear as loud what you hear as soft, and vice versa? We can see that this is impossible, since the two ends of the loudness scale have different properties; in particular, when loudness decreases to silence, all pitches become the same, but pitches retain their identity as loudness increases. Not only can we not hear different pitches of silence, we cannot even imagine them. Apparently our auditory systems do not have separate representations for 'silent middle C' and the 'silent A' above it; we can put the words together, but we cannot imagine the sounds. In other words, there is a 'degeneracy' or 'singularity' at zero-amplitude, where all pitches collapse together. As Francis Bacon said (*Essays*, 3), 'All colours will agree in the dark'; likewise, in silence all pitches are identical.

Next consider a pitch inversion, in which the sensation of high and low are reversed.[20] This is also impossible, because of the unique characteristics of low pitches: for if we listen to a sine wave of decreasing frequency, our perception of it will change gradually from a tone, to a buzz, to a rhythm. Neurologically, a pitch, which is mapped spatially in the auditory cortex, changes to an amplitude variation, which is mapped temporally. As the pitch decreases below about 1000 Hz., the nerve impulses begin to synchronize with the sound vibrations; below about 20 Hz., they are not perceived as pitch, but as periodic loudness variations (rhythm). In other words, at low frequencies the pitch and loudness axes are not independent; this does not happen at high frequencies, so the low end is differently structured topologically from the high end. The proposed inversion is impossible.

Finally, we may ask whether the pitch and loudness axes could be exchanged, so that I experience as pitch what you experience as loudness, and

[19] Thus I disagree with Chalmers (1995) when he asserts, 'There are properties of experience, such as the intrinsic nature of the sensation of red, that cannot be fully captured in a structural description'. I will argue that the experience is exhausted by its structure.

[20] The analysis here addresses pitch inversions rather than color inversions, since color vision is considerably more complex than hearing. Nevertheless, I am confident that a similar analysis will show the impossibility of a color inversion.

vice versa, but this is also impossible, because the two interact in an asymmetric way: low pitches grade into loudness variations, but soft sounds do not grade into pitch variations.

What may we conclude from the impossibility of these inversions? First, that subjective experience of sound must be just the way it is. For example the hearing of a low pitch is identical to intensification of certain pitch protophenomena that are connected in a certain way with loudness protophenomena. This view may seem tautologous, and therefore useless, but it is not. For example, if we discovered an organism with sense organs sensitive to vibrations of another kind (electrical, say), but of similar frequency, so that similar interrelations hold among the frequency and amplitude protophenomena, we could reasonably conclude that its experience of those sensations would be like sound. (Sensory qualities are explored in more detail in MacLennan, 1995.)

3.4 Unity of Consciousness

A phenomenal world derives its structure from the dependencies between protophenomena, which correspond to connections between activity sites; thus the unity of a phenomenal world is a consequence of this connectivity. We see this in split-brain operations (cerebral commissurotomy), wherein severing the corpus collosum causes a split in consciousness: each hemisphere is unconscious of what the other is experiencing (Gregory, 1987, pp. 740–7). However, it is significant that these operations do not completely separate the hemispheres; at very least the brainstem is left intact. Therefore the protophenomena corresponding to the two hemispheres are not completely independent, and so the phenomenal world has separated into two loosely-coupled subworlds.

An analogy may clarify this. A picture is an emergent effect of its individual pixels and their relative positions. If we cut a picture in half, it becomes two pictures, because there is no longer a fixed relations between pixels in one half and those in the other. However, instead of cutting the picture, we may gradually separate it into two parts, pixel by pixel, by stretching and eventually breaking the connections between them. The gradual uncoupling of the pixels in the two halves causes the picture to change gradually from one to two. So also, consciousness is emergent from the individual protophenomena and the dependencies between them. As the neural connections are weakened or broken, the protophenomena in the two subworlds decouple from each other, and the one mind becomes two.

This thought experiment demonstrates that the unity of consciousness is a matter of degree. Indeed, in principle we can measure the unity of consciousness by the tightness of the coupling between its protophenomena, for it is this coupling that gives the phenomenal world its coherence. (The tightness of coupling can, in principle, be calculated from the characteristic patterns; it can be quantified in terms of mutual information.)

One may wonder what sort of coupling is sufficient to unify consciousness. For example, in split-brain patients it has been observed that one hemisphere may communicate with the other through transactions with the external world, for example, twitching the skin on one side of the face so that it can be felt on the other. (The patient is unconscious of doing this.) Can 'external transactions' such as these effect the coupling of protophenomena? If so, then our individual phenomenal worlds may not be so independent as we commonly suppose, for any sort of communication couples protophenomena in one mind to those in another. I think the answer is, again, a matter of degree. There is an enormous difference between the bandwidth of the corpus collosum (approximately 800 million nerve fibers) and the narrow bandwidth of most external media. Nevertheless, the interconnection of phenomenal worlds by nonneural physical processes is a thought-provoking possibility.

3.5 The Unconscious Mind

The present view, which associates protophenomena — elementary units of consciousness — with all synapses, would seem to leave no room for the unconscious mind.[21] There are several possible resolutions.

(1) Unconscious processes may correspond to low-intensity, loosely-coupled protophenomena. By becoming coherent they come into consciousness (i.e. cohere into phenomena). That is, unconscious processes are incoherent pat-

[21] There are a number of definitions of the unconscious; for my purposes Jung's is as good as any: 'Everything of which I know, but of which I am not at the moment thinking; everything of which I was once conscious but have now forgotten; everything perceived by my senses, but not noted by my conscious mind; everything which, involuntarily and without paying attention to it, I feel, think, remember, want, and do; all future things that are taking shape in me and will sometime come into consciousness: all this is the content of the unconscious.' (CW 8, ¶382; Storr, p. 425)

terns in protophenomenal intensity. Therefore, unconscious processes are not literally unconscious; they are present in consciousness as a kind of background noise until and unless they cohere into macroscopic phenomena.

An analogy may clarify this. Project a slide on a screen, and defocus the lens. All of the same information is being projected on the screen as before, but now it is incoherent and the pattern is not salient; this is analogous to unconscious patterns in the protophenomena: they are there but not manifest. Focusing the lens makes the image manifest, which is analogous to the emergence of the unconscious content into conscious experience.

(2) The split-brain operations suggest another solution: in many cases the right hemisphere is unable to respond verbally to problems, and so it cannot easily manifest its consciousness to observers. Further, since the consciousness of the right hemisphere is largely disjoint from that of the left, the right forms a kind of unconscious mind for the left. Of course, the right hemisphere is as conscious as the left, and can manifest its consciousness in other ways, but its experience is not part of the left hemisphere's experience (or vice versa). The analogy becomes more striking when we recall that in these patients the hemispheres are not completely disconnected, so the right hemisphere can inject ideas into the left via the brainstem or via external transactions. Indeed, split-brain patients experience these communications as inexplicable 'hunches' — just like those from the unconscious (Gregory, 1987, p. 743).[22] In summary, what the perceiving-acting-speaking ego experiences as the 'unconscious mind' may be an equally conscious but loosely coupled part of the phenomenal world, which manifests itself only through hunches, dreams, urges, etc. More precisely, my phenomenal world may comprise two (or more) loosely coupled populations of tightly coupled protophenomena. One of these subworlds, which includes the motor protophenomena, is identified with the conscious ego because it can manifest its consciousness in behavior. However, other populations may be just as conscious, but unable to declare or demonstrate their consciousness to observers.

(3) Finally, according to the hypothesis of Sherrington and Pribram discussed earlier, consciousness is associated with graded dendritic microprocesses but not with all-or-none impulses in the axon. Therefore the uncon-

[22]This experiencing of interhemispheric communication as communications from an external source ('the gods') is of course a major premise of Jaynes' (1976) theory of the development of consciousness.

scious mind may reside in the axons, which would make it comprise the more reflexive or instinctive aspects of the psyche. In fact, such a model fits well with Jung's description of the unconscious, for he stressed that the 'archetypes of the collective unconscious' are contentless behavioral patterns grounded in our shared biological — or even physical — nature.[23] Thus they correspond to the axonal pathways, which are for the most part genetically determined. On the other hand, when an archetype emerges into consciousness, it does so with some individual content, which determines its particular appearance. The conscious manifestation of the archetypes corresponds to the dendritic microprocesses triggered by the axonal processes.

In fact, it is reasonable to suppose that all three of these explanations apply to the unconscious mind (which is primarily, it must be noted, a negative concept).

4 Comparison to Other Work

The philosophical view advocated here is consistent with that proposed independently by Chalmers (1995, 1996). First, it acknowledges his distinction between consciousness, the phenomenal world as experienced, and what he calls 'awareness', the neurological correlates of consciousness as an emergent physical phenomenon. Next, it is consistent with his 'principle of coherence', which postulates a direct correspondence between the structure of consciousness and the structure of awareness, since the phenomenological dependencies between protophenomena directly parallel the neurological dependencies between synapses, even to the extent of obeying the same mathematical laws (Appendix), so the emergence of higher level structures is also parallel. Third,

[23] "Again and again I encounter the mistaken notion that an archetype is determined in regard to its content, in other words, that it is a kind of unconscious idea (if such an expression be admissible). It is necessary to point out once more that archetypes are not determined as regards their content, but only as regards their form and then only to a very limited degree. A primordial image is determined as to its content only when it has become conscious and is therefore filled out with the material of conscious experience. ...The archetype in itself is empty and purely formal, nothing but a *facultas praeformandi*, a possibility of representation which is given *a priori*. The representations themselves are not inherited, only the forms, and in that respect they correspond in every way to the instincts, which are also determined in form only." (Jung, CW 9 i, ¶155; Storr, p. 415–6)

my theory is consistent with his 'principle of organizational invariance', which postulates that identity of microscopic functional organization implies qualitative identity of experience, since it is the dynamical interdependencies among protophenomena that create the phenomenal world. Finally, my view is compatible with his 'double-aspect principle', which hypothesizes that information has two aspects, one phenomenal and one physical. In the present theory, the basic units of information have a phenomenal aspect as protophenomena in consciousness and a physical aspect as activity sites in the brain. Beyond that, the theory of protophenomena is a step toward the sort of fundamental theory for which Chalmers has called, for it postulates a simple theoretical entity governed by mathematical laws, which provides a foundation for understanding the structure and dynamics of consciousness.

There are some superficial similarities between protophenomena and the *psychons* proposed by Sir John Eccles (1990, 1993); they are both elementary units of consciousness associated with synaptic activity in dendrites. The first difference is one of scale: Eccles associates psychons with *dendrons*, bundles of the apical dendrites of approximately one hundred pyramidal cells. Therefore, a dendron contains approximately 100,000 synapses, and so we could say that a psychon corresponds to approximately 100,000 protophenomena. The second difference is ontological, for Eccles' theory is explicitly dualistic. He takes a psychon to be a causal primary, which can influence synaptic processes by momentarily altering the quantum mechanical probability of an exocytosis of neurotransmitter into the synaptic cleft. In contrast, my theory is essentially monistic, for it views the phenomenal and the physical as two aspects of the same reality; in this sense my view is more akin to Sperry's non-dualist mentalism (Sperry, 1983, ch. 6).

On the other hand, Sperry's theory of *emergent causation* does not seem to adequately distinguish awareness and consciousness. Sperry (1983, p. 92) says, 'Once generated from neural events, the higher order mental patterns and programs have their own subjective qualities and progress, operate and interact by their own causal laws and principles which are different from, and cannot be reduced to those of neurophysiology....' The focus of emergent causation is on the functional role of conscious phenomena, and so he is dealing with awareness rather than consciousness; he holds 'subjective mental phenomena to be primary, causally potent realities as they are experienced subjectively' (p. 79), and speaks of the distinction between 'the causal potency of mental experience per se and that of its neural correlates' (p. 91), but does not address the emergence of subjective experience from smaller elements of subjectivity.

The theory presented in this paper has both philosophical and scientific aspects. As Sperry (1983, pp. 93, 99–103), Searle (1992, pp. 54–5) and others have noted, distinctions such as monism/dualism and mentalism/physicalism have outlived their usefulness, and their use to classify views such as ours are more likely to be misleading than helpful. Nevertheless, it is worthwhile to explain the philosophical aspects in these terms.

The present theory is dualistic in the sense that certain objects in certain situations (namely, activity sites in a functioning brain) have fundamental properties (protophenomena and their intensities), which are not reducible to physical properties. It is also dualistic in that the inherently private fact of experience is not reducible to the phenomena experienced, which are all potentially public (through a consensus of trained observers). Nevertheless, it is a kind of monism in postulating one 'stuff', which happens to have two fundamental, mutually irreducible aspects (phenomenal and physical).

Irreducibility enters in another way, for emergent causation operates in both the phenomenological (mental) and neurological (physical) realms: macroscopic consciousness governs microscopic protophenomenal dynamics (without violating the microscopic protophenomenal laws), as macroscopic awareness governs microscopic neurodynamics (without violating microscopic neurophysiology). (See also Sperry, 1983, pp. 93–6.) Once the philosophical arguments for irreducibility are granted, scientific investigation can proceed by parallel analyses in the phenomenological and neurological realms, each supplying the other with hypotheses, theories and empirical data. However, phenomenologically trained observers will be needed to obtain repeatable observations of the characteristics of consciousness.

5 Summary

As a first step I have proposed a theoretical entity, the protophenomenon, as an elementary unit of consciousness associated with microscopic activity sites in the brain, tentatively identified with the synapses. Like other theoretical entities in science, protophenomena are validated by their explanatory value and their fruitfulness for further progress. According to this theory the phenomenal world is structured by dynamical dependencies among the protophenomena, which parallel the neurodynamical dependencies among the corresponding activity sites; indeed they are described by the same mathematical laws.

6 References

Chalmers, D. J. (1995), 'Facing Up to the Problem of Consciousness', *Journal of Consciousness Studies*, **2**, pp. 200–219.

Chalmers, D. J. (1996), *The Conscious Mind* (New York: Oxford University Press).

Dennett, D. C. (1991), *Consciousness Explained* (Boston, Toronto and London: Little, Brown and Company).

Eccles, J. C. (1990), 'A Unitary Hypothesis of Mind-brain-interaction in the Cerebral Cortex', *Proceedings Royal Society London B*, **240**, pp. 433–451.

Eccles, J. C. (1993), 'Evolution of Complexity of the Brain with the Emergence of Consciousness', in *Rethinking Neural Networks: Quantum Fields and Biological Data*, ed. K. H. Pribram (Hillsdale: Lawrence Erlbaum).

Fleck, L. (1979), *Genesis and Development of a Scientific Fact*, transl. by Fred Bradley & Thaddeus J. Trenn (Chicago and London: University of Chicago Press).

Gregory, R. L. (ed.) (1987), *The Oxford Companion to the Mind* (Oxford and New York: Oxford University Press).

Hameroff, S. R. (1994), 'Quantum Coherence in Microtubules: A Neural Basis for Emergent Consciousness?', *Journal of Consciousness Studies*, **1**(1), pp. 91–118.

Hempel, C. G. (1965), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science* (New York: The Free Press).

Ihde, D. (1986), *Experimental Phenomenology: An Introduction* (Albany: State University of New York Press).

Jaynes, J. (1976), *The Origin of Consciousness in the Breakdown of the Bicameral Mind* (Boston: Houghton Mifflin).

MacLennan, B. J. (1991), 'Gabor Representations of Spatiotemporal Visual Images', Technical Report CS-91-144 (Knoxville: University of Tennessee Computer Science Department); accessible via URL `http://www.cs.utk.edu/~mclennan`.

MacLennan, B. J. (1993a), 'Grounding Analog Computers', *Think*, **2**, pp. 48–51.

MacLennan, B. J. (1993b), 'Information Processing in the Dendritic Net', in *Rethinking Neural Networks: Quantum Fields and Biological Data*, ed. K. H. Pribram (Hillsdale: Lawrence Erlbaum).

MacLennan, B. J. (1994), ' "Words Lie in Our Way" ', *Minds and Machines*, **2**(4), pp. 421–37.

MacLennan, B. J. (1995), 'The Investigation of Consciousness through Phenomenology and Neuroscience', in *Scale and Conscious Experience: Is the Brain Too Important to be Left to Specialists to Study?*, ed. J. King & K. H. Pribram (Hillsdale: Lawrence Erlbaum).

MacLennan, B. J. (in press), 'The Elements of Consciousness and their Neurodynamical Correlates', *Journal of Consciousness Studies*, in press.

Maxwell, G. (1980), 'The Ontological Status of Theoretical Entities', in *Introductory Readings in the Philosophy of Science*, ed. by E. D. Klemke, Robert Hollinger & A. David Kline (Buffalo: Prometheus Books).

Miller, G. A., Galanter, E., & Pribram, K. H. (1960), *Plans and the Structure of Behavior* (New York: Adams-Bannister-Cox).

Pribram, K. H. (1971), *Languages of the Brain: Experimental Paradoxes and Principles in Neuropsychology* (Englewood Cliffs: Prentice-Hall).

Pribram, K. H. (1991), *Brain and Perception: Holonomy and Structure in Figural Processing* (Hillsdale: Lawrence Erlbaum).

Sacks, O. (1985), *The Man who Mistook his Wife for a Hat* (London: Duckworth).

Sanger, T. D. (submitted), 'Probability Density Estimates for the Interpretation of Neural Population Codes', submitted to *Journal of Neurophysiology*.

Searle, J. (1992), *Rediscovery of the Mind* (Cambridge: MIT Press).

Shepherd, G. M. (1994), *Neurobiology*, 3rd ed. (New York and Oxford: Oxford University Press).

Sperry, R. (1983), *Science and Moral Priority: Merging Mind, Brain, and Human Values* (New York: Columbia University Press).

Storr, A. (1983), *The Essential Jung* (Princeton: Princeton University Press).

7 Appendix: Mathematical Model of Protophenomena

In this appendix I will outline a mathematical theory of the dynamical dependencies among protophenomena. The close relation between protophenomena and activity sites in the nervous system permits the equations governing the activity at those sites to be transferred to protophenomenal intensities. I will illustrate this transfer, based on the hypothesis that the activity sites are synapses and their activity is postsynaptic potential.

In terms of its electrical activity, a synapse is a voltage-controlled voltage source. On the presynaptic side it exhibits both resistance and capacitance, so we can treat a synapse as the equivalent circuit (Fig. 1). The dendrites, which connect many synapses, have resistance and capacitance, which depend on the diameter of the dendrite, among other things. As is commonly done, I will treat the dendrite as a tree of cylindrical segments of constant diameter.[24]

The dendritic membrane can be treated as a simple passive circuit, as shown in Fig. 2. Its impedance is $Z = R + 1/Cs$, since $1/Cs$ is the impedance

[24]Since we are primarily concerned with dendritic interactions (as the substrate of consciousness), I will not address the soma or axon here.

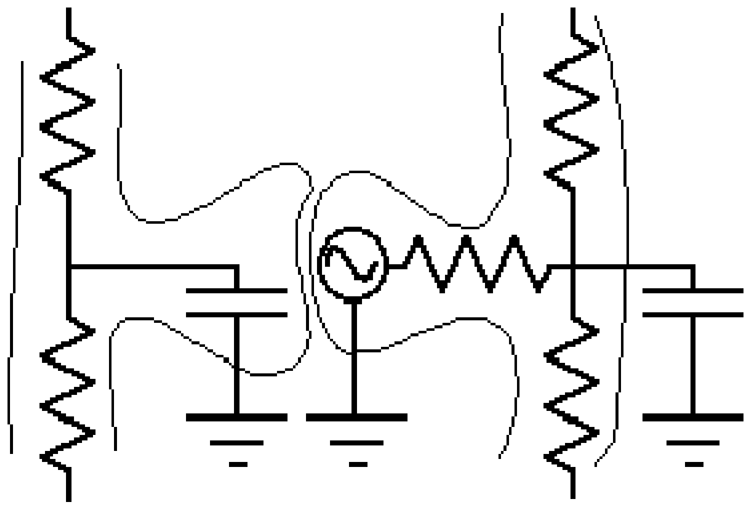

Figure 1: Equivalent circuit of a synapse

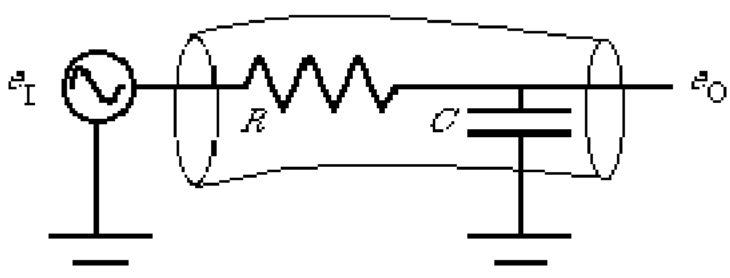

Figure 2: Equivalent circuit of a dendritic membrane

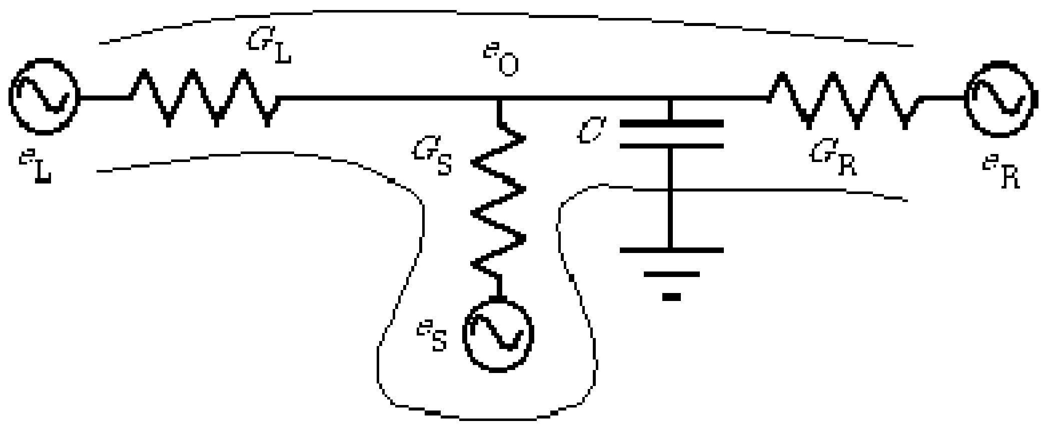

Figure 3: Equivalent circuit of postsynaptic membrane

of the membrane capacitance at frequency s. Thus the transfer function of the potential is given by

$$H(s) = \frac{1/Cs}{R + 1/Cs}.$$

Multiplying above and below by C and recalling that $\tau = RC$ is the time constant of the R-C circuit yields

$$H(s) = \frac{1}{\tau s + 1}.$$

The corresponding characteristic function (impulse response) is

$$h(t) = \exp(-t/\tau)/\tau.$$

This decaying exponential smoothes any impulse traversing the dendrite (with more smoothing for a larger time constant τ).

Consider now the effect of postsynaptic potential on dendritic membrane potential. Figure 3 shows the dendritic connections and the equivalent circuit. The output potential e_O at the root of the dendritic spine is an effect of the postsynaptic potential e_S, acting through the spine neck, interacting with the dendritic membrane potentials on both sides ($e_{\rm L}$ and $e_{\rm R}$). By applying Kirkhoff's laws we can determine $e_{\rm O}$ in terms of the dendritic conductances ($G_{\rm L}$, $G_{\rm S}$ and $G_{\rm R}$) and the membrane capacitance C near the spine. The dependence is easiest to express in terms of the Laplace transforms of the quantities:

$$E_{\rm O} = \frac{G_{\rm L}E_{\rm L} + G_{\rm S}E_{\rm S} + G_{\rm R}E_{\rm R}}{Cs + G_{\rm L} + G_{\rm S} + G_{\rm R}}.$$

Dividing above and below by C and recognizing that $G_{\rm x}/C = 1/CR_{\rm x} = 1/\tau_{\rm x}$ ($\rm x = L, S, R$), the inverse time constants of the dendritic segments, we have:

$$E_{\rm O} = \frac{E_{\rm L}/\tau_{\rm L} + E_{\rm S}/\tau_{\rm S} + E_{\rm R}/\tau_{\rm R}}{s + 1/\tau_{\rm L} + 1/\tau_{\rm S} + 1/\tau_{\rm R}}.$$

Let $u = 1/\tau_{\rm L} + 1/\tau_{\rm S} + 1/\tau_{\rm R}$ be the sum of the inverse time constants, and define the transfer function $H(s) = 1/(s+u)$. Then the Laplace transform of the output potential is

$$E_{\rm O} = H(s)(E_{\rm L}/\tau_{\rm L} + E_{\rm S}/\tau_{\rm S} + E_{\rm R}/\tau_{\rm R}).$$

The impulse function corresponding to H is $h(t) = \exp(-t/u)$. The output potential is then given by a convolution:

$$e_{\rm O}(t) = \exp(-t/u) \otimes [e_{\rm L}(t)/\tau_{\rm L} + e_{\rm S}(t)/\tau_{\rm S} + e_{\rm R}(t)/\tau_{\rm R}].$$

The weighted sum of the potentials is smoothed by the exponentially decreasing impulse function, a consequence of the membrane capacitance. This dependence between the potentials will be depicted by a symbol such as that in Fig. 4.

The presynaptic membrane potential $e_{\rm I}$ is a similar and in fact simpler convolution:

$$e_{\rm I}(t) = \exp(-t/u) \otimes [e_{\rm L}(t)/\tau_{\rm L} + e_{\rm R}(t)/\tau_{\rm R}],$$

where $u = 1/\tau_{\rm L} + 1/\tau_{\rm R}$. To a first approximation, the relation between the pre- and postsynaptic potentials is a simple proportion, $e_{\rm O} = ce_{\rm I}$. (To be more accurate we would have to consider the diffusion of the neurotransmitter, which would also have a smoothing effect.) I will depict this relationship by the symbol in Fig. 5. The characteristic pattern then is a vector function displaying the signals $(e_{\rm R}, e_{\rm L})$ to which the synapse is tuned:

$$\mathbf{h}_{\rm I}(t) = h(t)(1/\tau_{\rm L}, 1/\tau_{\rm R}) = [\exp(-t/u)/\tau_{\rm L}, \exp(-t/u)/\tau_{\rm R}].$$

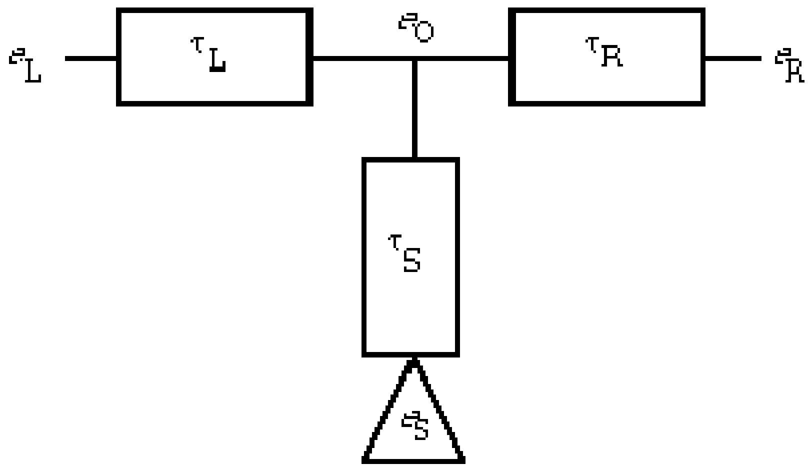

Figure 4: Diagram of dependence between postsynaptic potentials

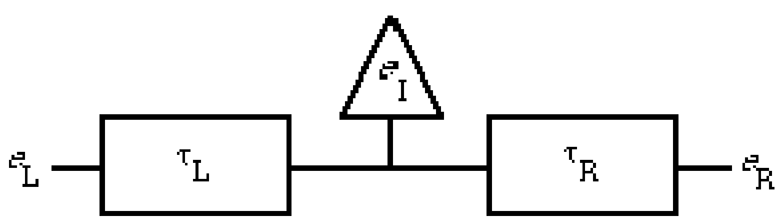

Figure 5: Diagram of dependence between presynaptic potentials

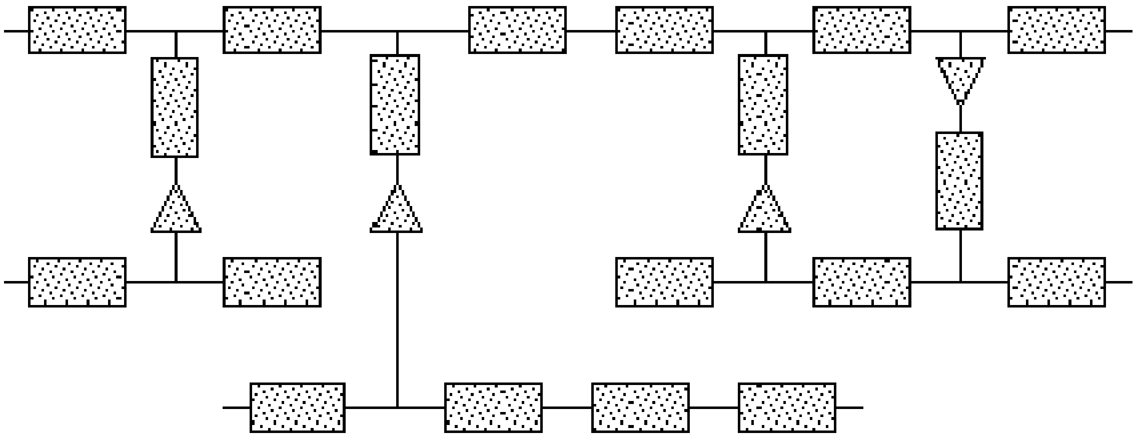

Figure 6: Example of dependencies between protophenomena

We now have formulas relating the activities at activity sites to each other, and we can reinterpret them as formulas relating the intensities of protophenomena. A diagram such as Fig. 6 shows how the protophenomenal intensities depend on each other, and allows their calculation, at least in principle. (It may be very complex in practice.) In general we can see that the (time-varying) intensity of a protophenomenon will be a complex function of the (time-varying) intensities of those on which it depends, as their intensities may be of its. The products of the transfer functions along each path to a synapse determine the transfer function of the synapse, and hence its characteristic pattern. The characteristic function of a protophenomenon is given by the same formula. The preceding analysis is based on an approximate linear model of passive dendritic processes; a more accurate analysis would have to take nonlinear effects into account.